dbGaP Study Release Notes

Release Notes for NHLBI TOPMed WGS GCPD-A, phs001661.v3.p1
"*NHLBI TOPMed: Genetic Causes of Complex Pediatric Disorders - Asthma (GCPD-A)*"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

October	19, 2020	Version 1 Data set release date
November	17, 2020	Version 2 Data set release date
August	31, 2021	Version 3 Data set release date

2021-08-31
Version 3 Data set release for NHLBI TOPMed WGS GCPD-A, now available

This release includes the addition of Freeze 9 whole genome sequences (WGS) and corresponding VCFs. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Disease-Specific (Asthma, GSO) (DS-ASTHMA-GSO)

Data Type	subjects	samples
Phenotype	5464	5464
Seq_DNA_SNP_CNV (VCFs)	4417	4417
WGS	4417	4417

For a description of SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Data Updates

1. **New Study Accession**

   NHLBI TOPMed WGS GCPD-A version 1 phs001661.v2.p1 has been updated to version 3. The dbGaP accession for the current set of data is **phs001661.v3.p1**. The participant number (p#) has not changed in version 3. No new subjects have been added to this study.

2. There are no updates to the phenotype datasets.

Molecular Data Updates

dbGaP QC steps for this release consist of checks for consistency of subject and sample IDs in phenotype and genotype components.

1. For samples and marker/enrichment-procedure info, see download components:
   a. phg001633.v1.TOPMed_WGS_GCPD_A_v3_frz9.sample-info.MULTI.tar.gz
2. Genotypes are available in a matrix format as multi-sample vcf file(s) packed within download component(s) marked as genotype-calls-vcf. Integrity of submitted vcf files and their compatibility with PSEQ are routinely checked. Components may be divided by platform and/or population.
   a. phg001633.v1.TOPMed_WGS_GCPD_A_v3_frz9.genotype-calls-vcf.WGS_markerset_grc38.c1.DS-ASTHMA-GSO.tar.gz

phg001462.v1	Freeze 8
phg001633.v1	Freeze 9

dbGaP Study Release Notes

Authorized Access (Individual Level Data)

Individual level data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001661/phs001661.v3.p1

2020-11-17

Version 2 Data set release for NHLBI TOPMed WGS GCPD-A, now available

This release includes the addition of the missing chrX VCF. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Disease-Specific (Asthma, GSO) (DS-ASTHMA-GSO)

Data Type	subjects	samples
Phenotype	5464	5464
Seq_DNA_SNP_CNV (VCFs)	4417	4417
WGS*	4417	4417

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.

For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Updates

1. **New Study Accession**

   NHLBI TOPMed WGS GCPD-A version 1 phs001661.v1.p1 has been updated to version 2. The dbGaP accession for the current phenotype data is **phs001661.v2.p1**. The participant number (p#) has not changed in version 2. No new subjects have been added to the study.

2. There are no changes to existing phenotype datasets.

Molecular Data Updates

phg001462 has been updated to v2. It includes the missing chrX.

1. For samples and marker/enrichment-procedure info, see download components:
   a. phg001462.v2.TOPMed_WGS_GCPD_A.sample-info.MULTI.tar.gz

   b. phg001462.v2.TOPMed_WGS_GCPD_A.marker-info.MULTI.tar.gz
2. Genotypes are available in a matrix format as multi-sample vcf file(s) packed within download component(s) marked as genotype-calls-vcf. Integrity of submitted vcf files and their compatibility with PSEQ are routinely checked. Components may be divided by platform and/or population.
   a. phg001462.v2.TOPMed_WGS_GCPD_A.genotype-calls-vcf.WGS_markerset_grc38.c1.DS-ASTHMA-GSO.tar.gz

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001661/phs001661.v2.p1

2020-10-19

Version 1 Data set release for NHLBI TOPMed WGS GCPD-A, now available

This release includes the addition of Freeze 8 whole genome sequences (WGS) brokered through the Sequence Read Archive (SRA), and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Disease-Specific (Asthma, GSO) (DS-ASTHMA-GSO)

Data Type	subjects	samples
Phenotype	5464	5464
Seq_DNA_SNP_CNV (VCFs)	4417	4417
WGS*	4417	4417

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.

For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Molecular Data

dbGaP QC steps for this release consist of checks for consistency of subject and sample IDs in phenotype and genotype components.

1. For samples and marker/enrichment-procedure info, see download components:
   a. phg001462.v1.TOPMed_WGS_GCPD_A.sample-info.MULTI.tar.gz

   b. phg001462.v1.TOPMed_WGS_GCPD_A.marker-info.MULTI.tar.gz
2. Genotypes are available in a matrix format as multi-sample vcf file(s) packed within download component(s) marked as genotype-calls-vcf. Integrity of submitted vcf files and their compatibility with PSEQ are routinely checked. Components may be divided by platform and/or population.
   a. phg001462.v1.TOPMed_WGS_GCPD_A.genotype-calls-vcf.WGS_markerset_grc38.c1.DS-ASTHMA-GSO.tar.gz

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs001661/phs001661.v1.p1